W372–W376 *Nucleic Acids Research, 2008, Vol. 36, Web Server issue* *Published online 8 May 2008*
doi:10.1093/nar/gkn252

The Ontology Lookup Service: more data and better tools for controlled vocabulary queries

Richard G. Côté*, Philip Jones, Lennart Martens, Rolf Apweiler and Henning Hermjakob

EMBL Outstation, European Bioinformatics Institute (EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

Received January 31, 2008; Revised April 10, 2008; Accepted April 20, 2008

ABSTRACT

The Ontology Lookup Service (OLS) (http://www.ebi.ac.uk/ols) provides interactive and programmatic interfaces to query, browse and navigate an ever increasing number of biomedical ontologies and controlled vocabularies. The volume of data available for querying has more than quadrupled since it went into production and OLS functionality has been integrated into several high-usage databases and data entry tools. Improvements have been made to both OLS query interfaces, based on user feedback and requirements, to improve usability and service interoperability and provide novel ways to perform queries.

INTRODUCTION

The current trend towards multi-domain data analysis, compounded with the vast amounts of data being generated by high-throughput methods, presents a significant data management challenge. Controlled vocabularies and ontologies therefore become crucial tools for data annotation and analysis to help provide a stable and consistent context for large data sets.

The Ontology Lookup Service (OLS) was created to provide a simple, centralized, integrated interface to query multiple biomedical ontologies by interactive and programmatic means. Prior to its creation, users wishing to query ontologies had to go to individual websites—when available—and use whatever query interface was made available. Many ontologies were only available in flat-file format and few ontologies could be queried by programmatic means. Querying multiple ontologies was a difficult and time-consuming proposition.

The OLS has been in production since mid-2005 and has proven to be a popular tool with data producers and consumers. The OLS has been previously described and readers are invited to refer to the original publication for in-depth information on the technical architecture and data models (1).

The core functionality of the OLS provides users with the means to perform queries on controlled vocabulary and ontology terms and synonyms, as well as navigate the relationships between terms and obtain additional metadata (such as definitions, comments, synonyms or cross-references to other databases) and annotations on selected terms.

Data producers can use the online interface to search for appropriate terms to annotate their submissions in one specific ontology or across all available ontologies served by the OLS. An ontology browser is also available to navigate ontologies and controlled vocabularies to find the appropriate context and level of detail for a given term. Application developers can use the provided web service interface to fully integrate OLS functionality within their applications.

Since its inception, it has become an integral part of many highly accessed databases (2–5) and has been recommended as a data submission resource by the Human Proteome Organization (HUPO) Proteomics Standards Initiative (PSI) (6) and by the International Molecular Interaction Exchange consortium (IMEX) (7).

Application developers are making use of the OLS web service interface to enrich their own applications. Examples of this include the Proteome Harvest data submission tool for PRIDE (8) and the Map2OWL Protégé plug-in (http://map2owl.sourceforge.net/).

This article describes the new data available in the OLS as well as the many improvements to both the online and the programmatic interfaces through which this data can be browsed and queried.

AVAILABLE DATA

When it first became publicly available, the OLS contained 42 ontologies, which accounted for close to 135 000 terms. Over a 2-year period, the data content of the OLS has grown to 58 ontologies and more than 595 000 terms (Figure 1). These cover far-ranging topics such as model organism anatomy and development, physiology and disease, instrumentation and methods and many others. Significant milestones for the OLS include the

*To whom correspondence should be addressed. Tel: + 44 1223 492 610; Fax: + 44 1223 494 468; Email: rcote@ebi.ac.uk

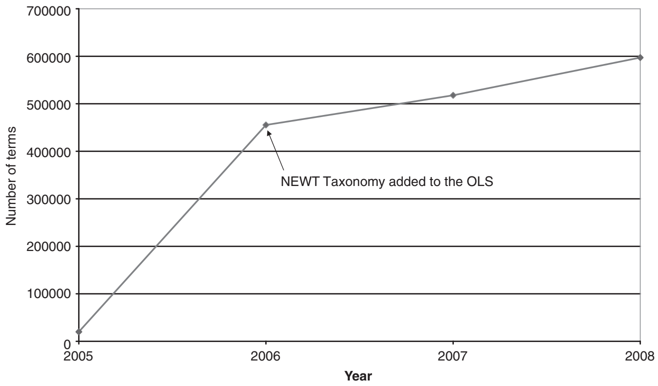


Figure 1. Growth chart of the OLS data content. The amount of data loaded into the OLS, based on unique terms, has more than quadrupled since the service went online. Note that the large increase in 2006 is largely due to the incorporation of the NEWT taxonomy.

incorporation of the NEWT taxonomy that provides information on over 400 000 taxonomic classifications (roughly 330 000 of which are species names) and the incorporation of all of the HUPO PSI domain ontologies.

Users are encouraged to go online at http://www.ebi.ac.uk/ontology-lookup/ontologyList.do to access a full listing of currently available ontologies and controlled vocabularies. The ontologies are created and maintained by experts in their respective field (4,9–12) and many of the more commonly used ones are mirrored by the Open Biomedical Ontology (OBO) Foundry (13). In order to provide the latest versions of these ontologies and maintain the OLS as up-to-date as possible, the ontology providers are polled on a daily basis and updated files are downloaded and parsed to update the core OLS database. The OLS data loaders have been run 835 times and have refreshed 1900 ontologies since September 1, 2006.

The OLS now provides a complete database export in MySQL format for users who wish a local copy of the relational data for their own queries. This database export is done on a weekly basis and can be obtained from the EBI public FTP server (ftp://ftp.ebi.ac.uk/pub/databases/ols). Information on how to use this database export can be found online at http://www.ebi.ac.uk/ontology-lookup/databaseExport.do.

INTERACTIVE USER INTERFACE IMPROVEMENTS

The OLS provides a rich browsing experience using AJAX technologies. A suggest-as-you type search mechanism has received very positive feedback from users who are looking for terms to annotate their data but are unsure where to start looking. Once a term is selected, metadata (definition, synonyms and cross-references) are fetched from the database and displayed to the user. Users can browse full ontologies or subsets of them with a click of a button. A graph of all possible paths from a selected term to the root of the ontology will be displayed (Figure 2).

Several improvements have been incorporated into the online interface. The first improvement is the possibility to include or exclude obsolete terms from the suggestion list by simply toggling a checkbox (located above the main search box). By default, terms that have been marked as obsolete by the ontology maintainers will be returned as suggested search results. Unchecking the box will prevent such terms from being returned by the system.

Another improvement has been the possibility to query the OLS by term identifier (e.g. GO:0008150) and to provide direct search URLs (e.g. http://www.ebi.ac.uk/ontology-lookup/?termId = GO:0008150) This feature allows users to provide links directly to specific terms, where the preferred term name and all known metadata are displayed.

A final improvement added to the user interface is the ability to browse the OLS loader statistics. A link from the statistics box on the main page allows users to see which ontologies have been recently updated as well as the number of terms loaded for each ontology. Users can navigate in monthly increments to obtain the relevant information.

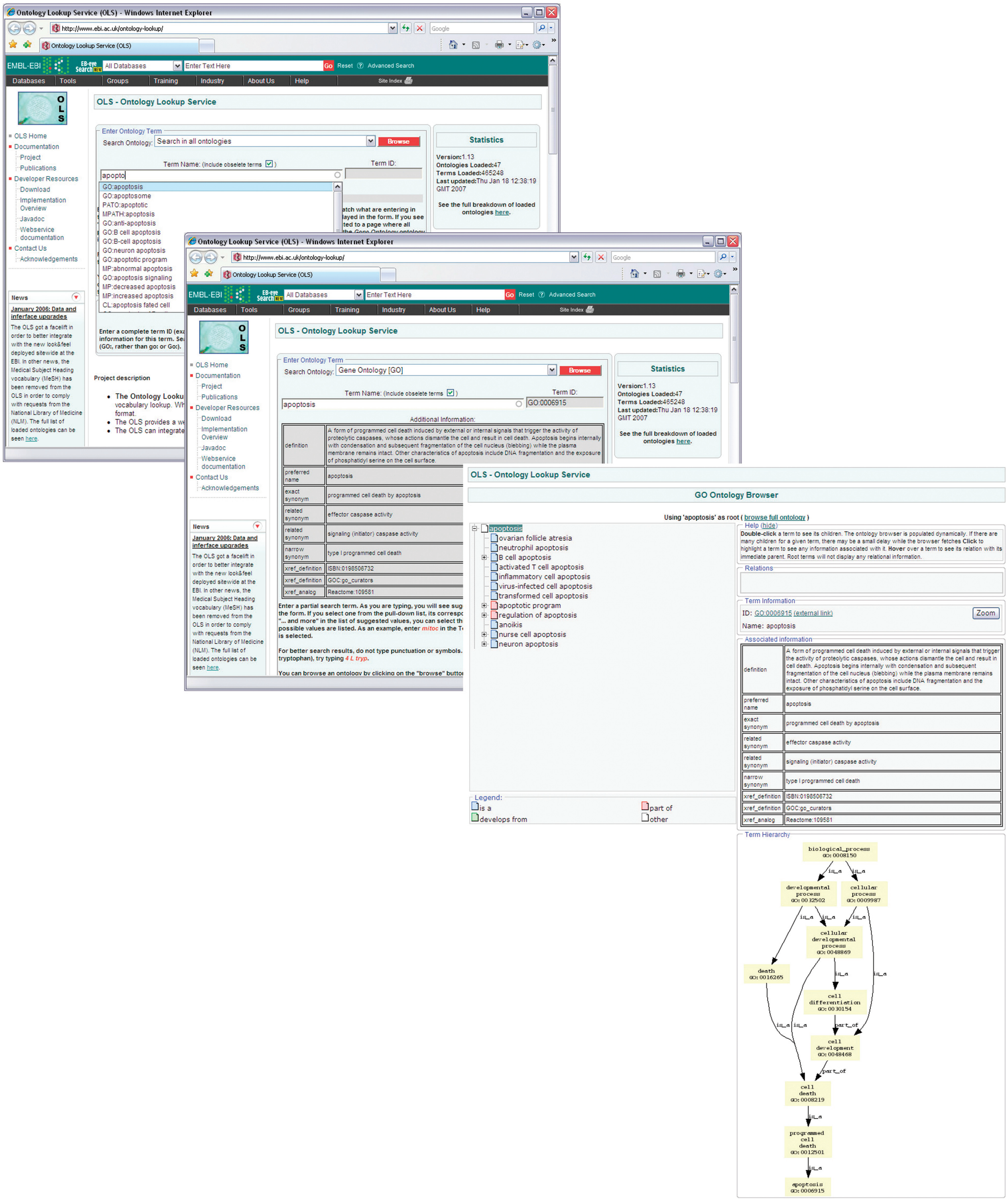
SOAP USER INTERFACE IMPROVEMENTS

The OLS allows application developers to query and retrieve data using a web service interface implemented using the Apache AXIS SOAP toolkit. The web service interface is described in a WSDL document that can be used by various high-level programming languages to properly create messages between the OLS server and the client application.

One of the strengths of SOAP is that it can be platform independent across multiple programming languages (e.g. client code written in Perl running on an Apple Mac OSX can communicate with a server written in Java running on Linux). This interoperability is not without its caveats, however, and there are implementation limitations because not all programming languages implement the full specification and not all implementations are equally good.

To mitigate these effects and improve interoperability across platforms, the WSDL document describing the OLS web service was recoded from ‘RPC/encoded’ to ‘document/literal’. These conventions dictate how to translate a WSDL binding to a SOAP message that can be exchanged between the client and the server and it is generally accepted that the ‘document/literal’ provides the highest possibility of interoperability. Furthermore, the original interface had overloaded method signatures, where the same method name had multiple argument lists. This proved to be problematic for certain platforms and the issue was resolved by giving each method a unique name. These improvements allowed the OLS to become usable in workflow engines such as Taverna (14).

The web service interface was also enriched to mirror work done on the interactive interface and allow a greater scope of information to be retrieved programmatically. New methods now allow users to obtain information on database cross-references and annotations and determine if terms are obsolete or active. Other methods provide easier means to navigate relations between terms in multiple directions (for example, obtaining all the child terms of a parent, or all the parents for a given child term).



Ontology Lookup Service (OLS) - Windows Internet Explorer

http://www.ebi.ac.uk/ontology-lookup/

OLS - Ontology Lookup Service

Enter Ontology Term

Search Ontology: Search in all ontologies

Term Name: (Include obselete terms ☑)

apopto

Term ID:

- GO:apoptosis
- GO:apoptosome
- PATO:apoptotic
- MPATH:apoptosis
- GO:anti-apoptosis
- GO:B cell apoptosis
- GO:B-cell apoptosis
- GO:neuron apoptosis
- GO:apoptotic program
- MP:abnormal apoptosis
- GO:apoptosis signaling
- MP:decreased apoptosis
- MP:increased apoptosis
- CL:apoptosis fated cell

atch what are entering in
layed in the form. If you see
ted to a page where all
the Gene Ontology ontology

Enter a complete term ID (exa
information for this term. Sea
(GO:, rather than go: or Go:).

Statistics

Version:1.13
Ontologies Loaded:47
Terms Loaded:465248
Last updated:Thu Jan 18 12:38:19 GMT 2007

See the full breakdown of loaded ontologies here.

News

January 2006: Data and inferface upgrades

The OLS got a facelift in order to better integrate with the new look&feel deployed sitewide at the EBI. In other news, the Medical Subject Heading vocabulary (MeSH) has been removed from the OLS in order to comply with requests from the National Library of Medicine (NLM). The full list of loaded ontologies can be seen here.

Project description

- The Ontology Looku
  vocabulary lookup. Wh
  format.
- The OLS provides a we
- The OLS can integrate

Ontology Lookup Service (OLS) - Windows Internet Explorer

http://www.ebi.ac.uk/ontology-lookup/

OLS - Ontology Lookup Service

Enter Ontology Term

Search Ontology: Gene Ontology [GO]

Term Name: (Include obselete terms ☑)

apoptosis

Term ID: GO:0006915

Additional Information:

definition	A form of programmed cell death induced by external or internal signals that trigger the activity of proteolytic caspases, whose actions dismantle the cell and result in cell death. Apoptosis begins internally with condensation and subsequent fragmentation of the cell nucleus (blebbing) while the plasma membrane remains intact. Other characteristics of apoptosis include DNA fragmentation and the exposure of phosphatidyl serine on the cell surface.
preferred name	apoptosis
exact synonym	programmed cell death by apoptosis
related synonym	effector caspase activity
related synonym	signaling (initiator) caspase activity
narrow synonym	type I programmed cell death
xref_definition	ISBN:0198506732
xref_definition	GOC:go_curators
xref_analog	Reactome:109581

Enter a partial search term. As you are typing, you will see sug
the form. If you select one from the pull-down list, its correspo
"... and more" in the list of suggested values, you can select th
possible values are listed. As an example, enter *mitoc* in the Te
is selected.

For better search results, do not type punctuation or symbols.
tryptophan), try typing *4 L tryp*.

You can browse an ontology by clicking on the "browse" button

Statistics

Version:1.13
Ontologies Loaded:47
Terms Loaded:465248
Last updated:Thu Jan 18 12:38:19 GMT 2007

See the full breakdown of loaded ontologies here.

OLS - Ontology Lookup Service

GO Ontology Browser

Using 'apoptosis' as root (browse full ontology)

- apoptosis
  - ovarian follicle atresia
  - neutrophil apoptosis
  - B cell apoptosis
  - activated T cell apoptosis
  - inflammatory cell apoptosis
  - virus-infected cell apoptosis
  - transformed cell apoptosis
  - apoptotic program
  - regulation of apoptosis
  - anoikis
  - nurse cell apoptosis
  - neuron apoptosis

Legend:

🟦 is a	🟥 part of
🟩 develops from	□ other

Help (hide)

Double-click a term to see its children. The ontology browser is populated dynamically. If there are many children for a given term, there may be a small delay while the browser fetches **Click** to highlight a term to see any information associated with it. **Hover** over a term to see its relation with its immediate parent. Root terms will not display any relational information.

Relations

Term Information

ID: GO:0006915 (external link)

Name: apoptosis

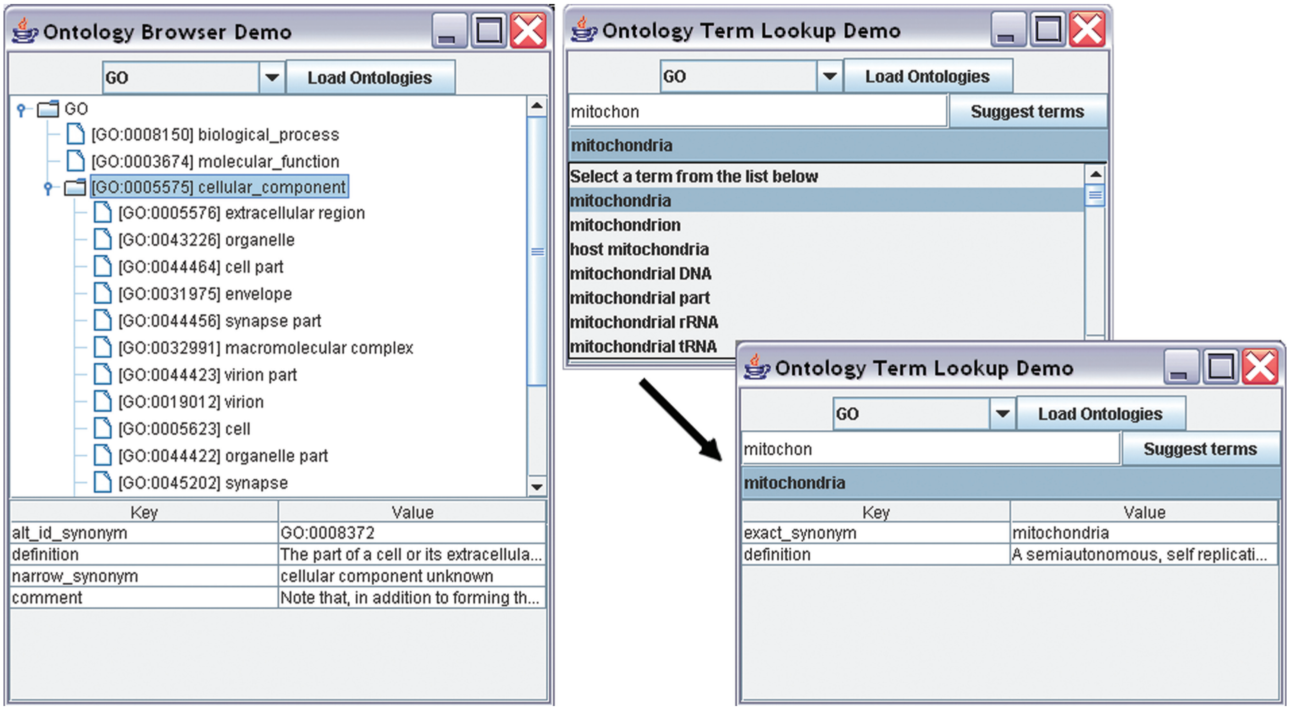
Associated information

definition	A form of programmed cell death induced by external or internal signals that trigger the activity of proteolytic caspases, whose actions dismantle the cell and result in cell death. Apoptosis begins internally with condensation and subsequent fragmentation of the cell nucleus (blebbing) while the plasma membrane remains intact. Other characteristics of apoptosis include DNA fragmentation and the exposure of phosphatidyl serine on the cell surface.
preferred name	apoptosis
exact synonym	programmed cell death by apoptosis
related synonym	effector caspase activity
related synonym	signaling (initiator) caspase activity
narrow synonym	type I programmed cell death
xref_definition	ISBN:0198506732
xref_definition	GOC:go_curators
xref_analog	Reactome:109581

Term Hierarchy

Figure 2. The interactive OLS user interface. A suggest-as-you-type search mechanism provides users with interactive term lists based on their input. Once a term is selected, known metadata are obtained and displayed. Users can browse an ontology using the selected term as a starting point. Using the ontology browser, term relations can be navigated. Highlighting a term will load metadata and provide a graphical display of all paths from that term to the ontology root(s).

Finally, ontology-level methods are also available: it is now possible to query the load date of an ontology and also obtain all the terms for an ontology in a single request. Please refer to the OLS online web service documentation for a complete technical review on available methods and how to use them. The documentation can be accessed at the following link: http://www.ebi.ac.uk/ontology-lookup/WSDLDocumentation.do.



Ontology Browser Demo — GO

```
GO
├ [GO:0008150] biological_process
├ [GO:0003674] molecular_function
└ [GO:0005575] cellular_component
   ├ [GO:0005576] extracellular region
   ├ [GO:0043226] organelle
   ├ [GO:0044464] cell part
   ├ [GO:0031975] envelope
   ├ [GO:0044456] synapse part
   ├ [GO:0032991] macromolecular complex
   ├ [GO:0044423] virion part
   ├ [GO:0019012] virion
   ├ [GO:0005623] cell
   ├ [GO:0044422] organelle part
   ├ [GO:0045202] synapse
```

Key	Value
alt_id_synonym	GO:0008372
definition	The part of a cell or its extracellula...
narrow_synonym	cellular component unknown
comment	Note that, in addition to forming th...

Ontology Term Lookup Demo — GO; mitochon; mitochondria

Select a term from the list below
mitochondria
mitochondrion
host mitochondria
mitochondrial DNA
mitochondrial part
mitochondrial rRNA
mitochondrial tRNA

Ontology Term Lookup Demo — GO; mitochon; mitochondria

Key	Value
exact_synonym	mitochondria
definition	A semiautonomous, self replicati...

Figure 3. Two Java applications using the SOAP interface. An ontology browser demo application and a term search demo application can be downloaded from the OLS website to illustrate the ease with which OLS functionality can be integrated within existing applications.

In order to showcase the ease in which the OLS can be integrated into existing applications, fully functional demonstration Java applications are made available from the 'download' section of the OLS website (Figure 3). These applications require Java 1.4 or later to run and include the complete source code that illustrates how to use Java to query the OLS web service and use the results.

DISCUSSION

The OLS has proven to be successful beyond its originally intended scope. Several projects, such as the HUPO PSI, BioSapiens and PRIDE use the OLS to host their own domain-specific ontologies and use it as their primary ontology browser. PRIDE and IntAct, among other projects, have successfully incorporated OLS functionality into their applications to enrich their query and data annotation interfaces.

The OLS is still under active development and ongoing work is currently in progress to bring the OLS web service in line with the latest web service specifications (WS-I compliance) and provide a richer object model for programmatic queries. New ontologies are always being added to the core database, either coming from the OBO Foundry or from direct user submissions. Usage statistics indicate that both the interactive and programmatic interfaces are showing ever increasing usage. Monthly usage has rapidly climbed from 120 000 hits in mid-2005 to over 700 000 hits by late 2007.

OLS development is highly driven by user requirements. Based on obtained feedback, updates to the online interface have already been implemented, as have been extensive upgrades to the web server interface. All of these, combined with complete code samples and FTP access to a complete database export, provide simple yet powerful methods to access ontology and controlled vocabulary data that should suit every user requirement.

ACKNOWLEDGEMENTS

OLS contributors are supported through the Biotechnology and Biological Sciences Research Council (BBSRC) ISPIDER grant and EU FP6 'Felics' [contract number 021902 (RII3)] grants. We would like to thank all users who have contributed recommendations and requests. Many improvements, already executed and some ongoing, come directly from this feedback. Funding to pay the Open Access publication charges for this article was provided by EU FP6 Felics [contract number 021902 (RII3)] grant.

Conflict of interest statement: None declared.

REFERENCES

1. Côté,R.G., Jones,P., Apweiler,R. and Hermjakob,H. (2006) The ontology lookup service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics*, **7**, 97.
2. Kerrien,S., Alam-Faruque,Y., Aranda,B., Bancarz,I., Bridge,A., Derow,C., Dimmer,E., Feuermann,M., Friedrichsen,A., Huntley,R. *et al.* (2007) Intact–open source resource for molecular interaction data. *Nucleic Acids Res.*, **35**, D561–D565.
3. Jones,P., Côté,R.G., Cho,S.Y., Klie,S., Martens,L., Quinn,A.F., Thorneycroft,D. and Hermjakob,H. (2008) Pride: new developments and new datasets. *Nucleic Acids Res.*, **36**, D878–D883.
4. Degtyarenko,K., de Matos,P., Ennis,M., Hastings,J., Zbinden,M., McNaught,A., Alcántara,R., Darsow,M., Guedj,M. and

Ashburner,M. (2008) Chebi: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res.*, **36**, D344–D350.

5. Mallon,A., Blake,A. and Hancock,J.M. (2008) Europhenome and empress: online mouse phenotyping resource. *Nucleic Acids Res.*, **36**, D715–D718.
6. Hermjakob,H. (2006) The HUPO proteomics standards initiative – overcoming the fragmentation of proteomics data. *Proteomics*, **6**, 34–38.
7. Orchard,S., Salwinski,L., Kerrien,S., Montecchi-Palazzi,L., Oesterheld,M., Stümpflen,V., Ceol,A., Chatr-Aryamontri,A., Armstrong,J., Woollard,P. *et al.* (2007) The minimum information required for reporting a molecular interaction experiment (mimix). *Nat. Biotechnol.*, **25**, 894–898.
8. Jones,P., Côté,R.G., Cho,S.Y., Klie,S., Martens,L., Quinn,A.F., Thorneycroft,D. and Hermjakob,H. (2008) Pride: new developments and new datasets. *Nucleic Acids Res.*, **36(Database issue)**, D878–D883.
9. Wilson,R.J., Goodman,J.L. and Strelets,V.B. (2008) Flybase: integration and improvements to query tools. *Nucleic Acids Res.*, **36**, D588–D593.
10. Swarbreck,D., Wilks,C., Lamesch,P., Berardini,T.Z., Garcia-Hernandez,M., Foerster,H., Li,D., Meyer,T., Muller,R., Ploetz,L. *et al.* (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.*, **36**, D1009–D1014.
11. Liang,C., Jaiswal,P., Hebbard,C., Avraham,S., Buckler,E.S., Casstevens,T., Hurwitz,B., McCouch,S., Ni,J., Pujar,A. *et al.* (2008) Gramene: a growing plant comparative genomics resource. *Nucleic Acids Res.*, **36**, D947–D953.
12. The Gene Ontology Consortium (2008) The gene ontology project in 2008. *Nucleic Acids Res.*, **36**, D440–D444.
13. Smith,B., Ashburner,M., Rosse,C., Bard,J., Bug,W., Ceusters,W., Goldberg,L.J., Eilbeck,K., Ireland,A., Mungall,C.J. *et al.* (2007) The OBO foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.*, **25**, 1251–1255.
14. Hull,D., Wolstencroft,K., Stevens,R., Goble,C., Pocock,M.R., Li,P. and Oinn,T. (2006) Taverna: a tool for building and running workflows of services. *Nucleic Acids Res.*, **34**, W729–W732.